dat*EAU*base: A powerful large capacity database for raw and validated water quality data with emphasis on their metadata

"EAU" is water in French

Q. Plana, T. Kraft, J. Alferes, T. Maruéjouls, P.A. Vanrolleghem

National Monitoring Conference, Tampa Bay
May 5, 2016
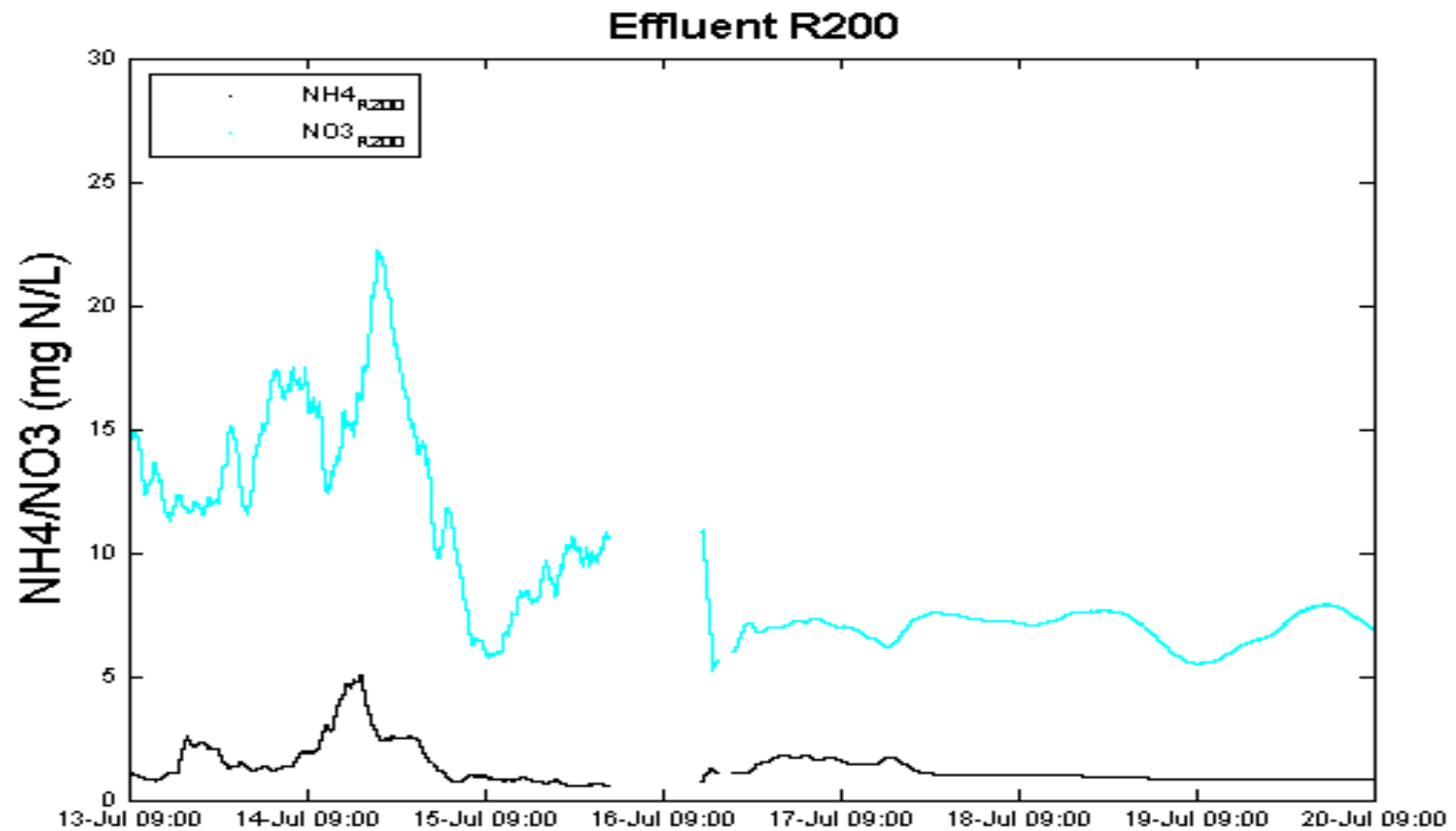
UNIVERSITÉ LAVAL

model*EAU*

Contents

- Introduction

- Types of data

- Challenges

- dat*EAU*base scheme

- dat*EAU*base design

- Examples of data

- Conclusions

Automated monitoring stations + Sampling/lab

Housing

Station

Sensors

Automated monitoring stations

High frequency monitoring

Large amounts of relevant data

Raw and filtered data

Importance of the metadata (data about data)

• What does this single value mean?

Importance of the metadata

- What does this single value mean?

What has been measured?	When has the value been measured?
Where has the value been measured?	7.2811	Why has the value been measured?
Who has collected the value?	How has the value been measured?

Importance of the metadata

- Collected data are only useful when:

  - They are well-documented

  - Their quality is assured

- Data storage task

- Data validation task

Contents
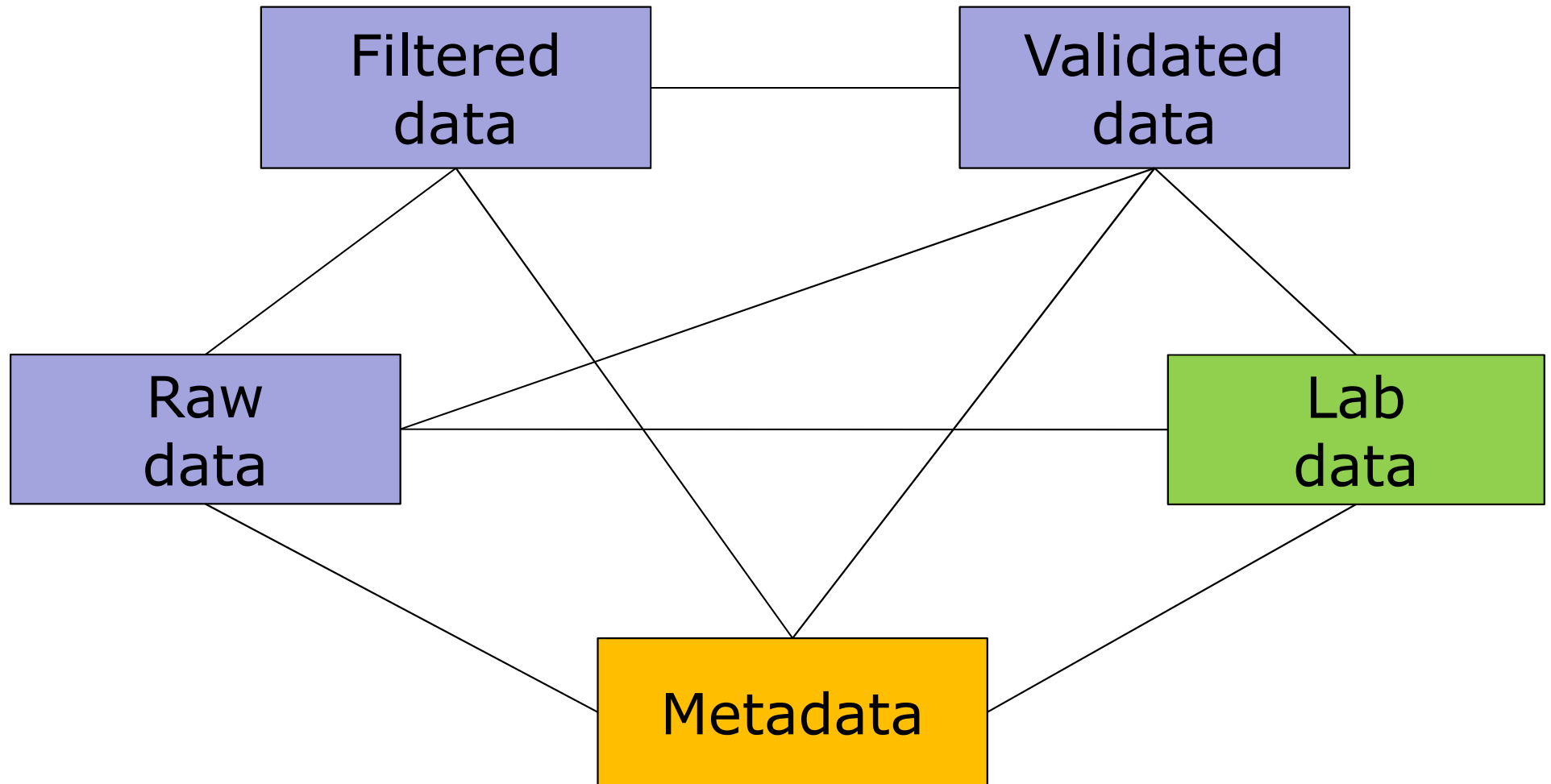
Types of data

Contents

Data management challenges

• Variability of raw data formats

• Databases grow continuously
  and monitoring programs change

      ➔ Database must be adapted

• Need for high-performance storage and data access

• Personnel that is collecting and managing data
  is changing over time

      ➔ Inconsistencies

Contents

dat*EAU*base scheme

MySQL

Database → Raw data

Filtered data — Validated data

Lab data

Metadata

Python

Graphical user interface ↔ Data import / Data overview / Data export

User
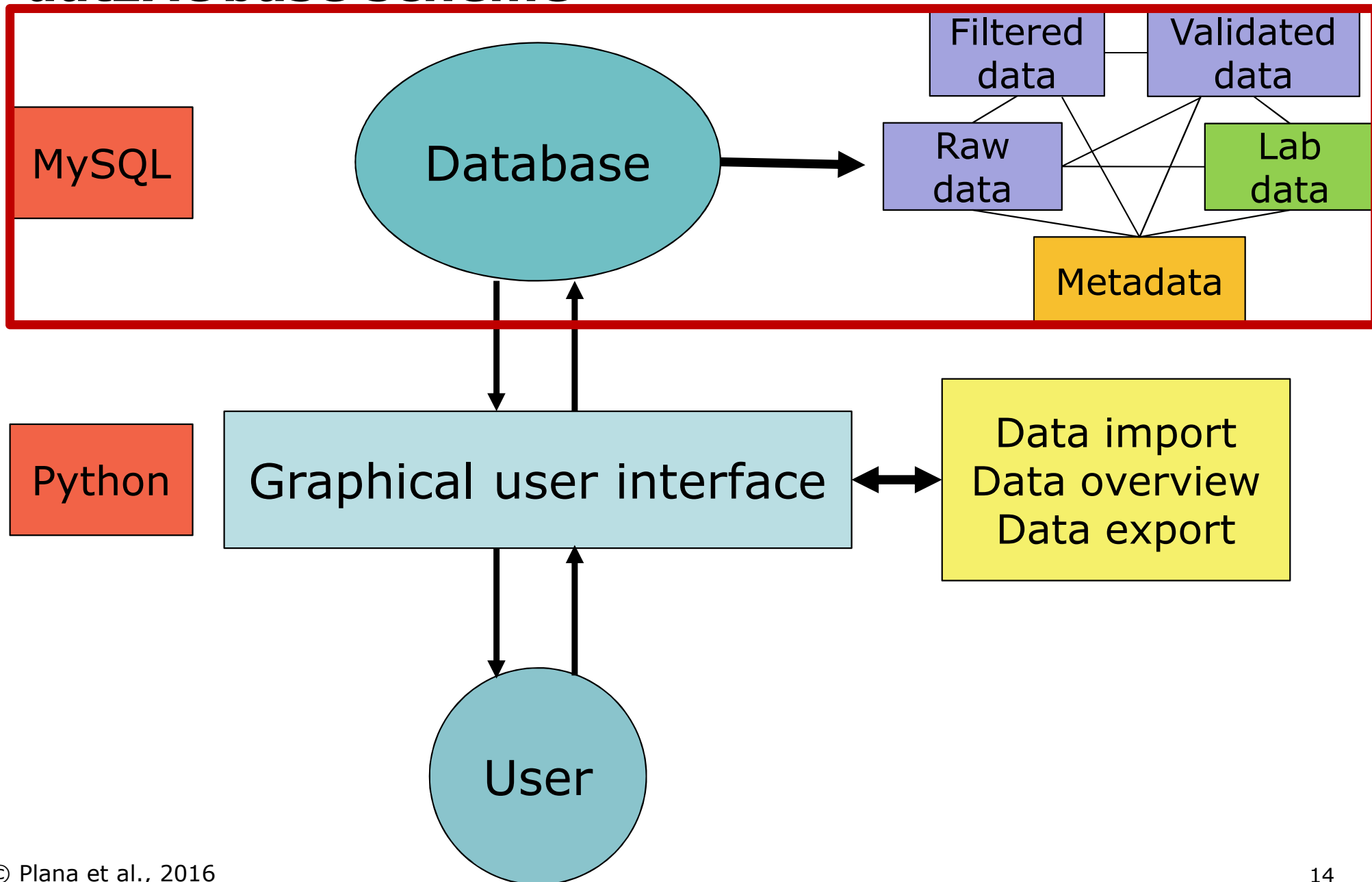
Contents

- Introduction

- Types of data

- Challenges

- dat*EAU*base scheme

- **dat*EAU*base design**

- Examples of data

- Conclusions

dat*EAU*base design

Purpose — Why?

Parameter — What?

Procedure

Equipment — How?

Site

Sampling Point

Watershed — Where?

Water related values

Contact — Who?

Weather conditions — When?

Project — Why?

16

dat*EAU*base design

Primary tables

Lookup tables

© Plana et al., 2016

17
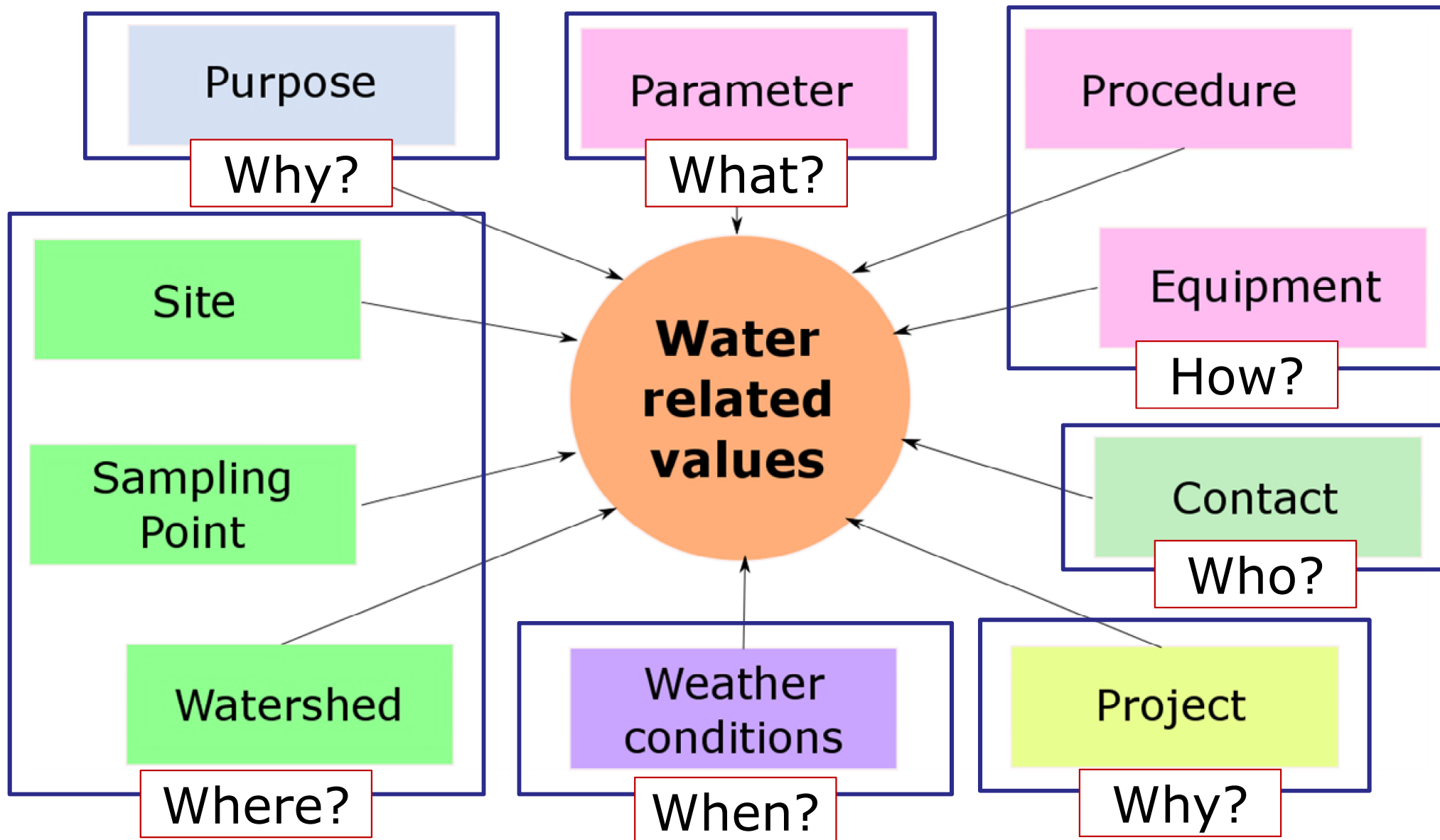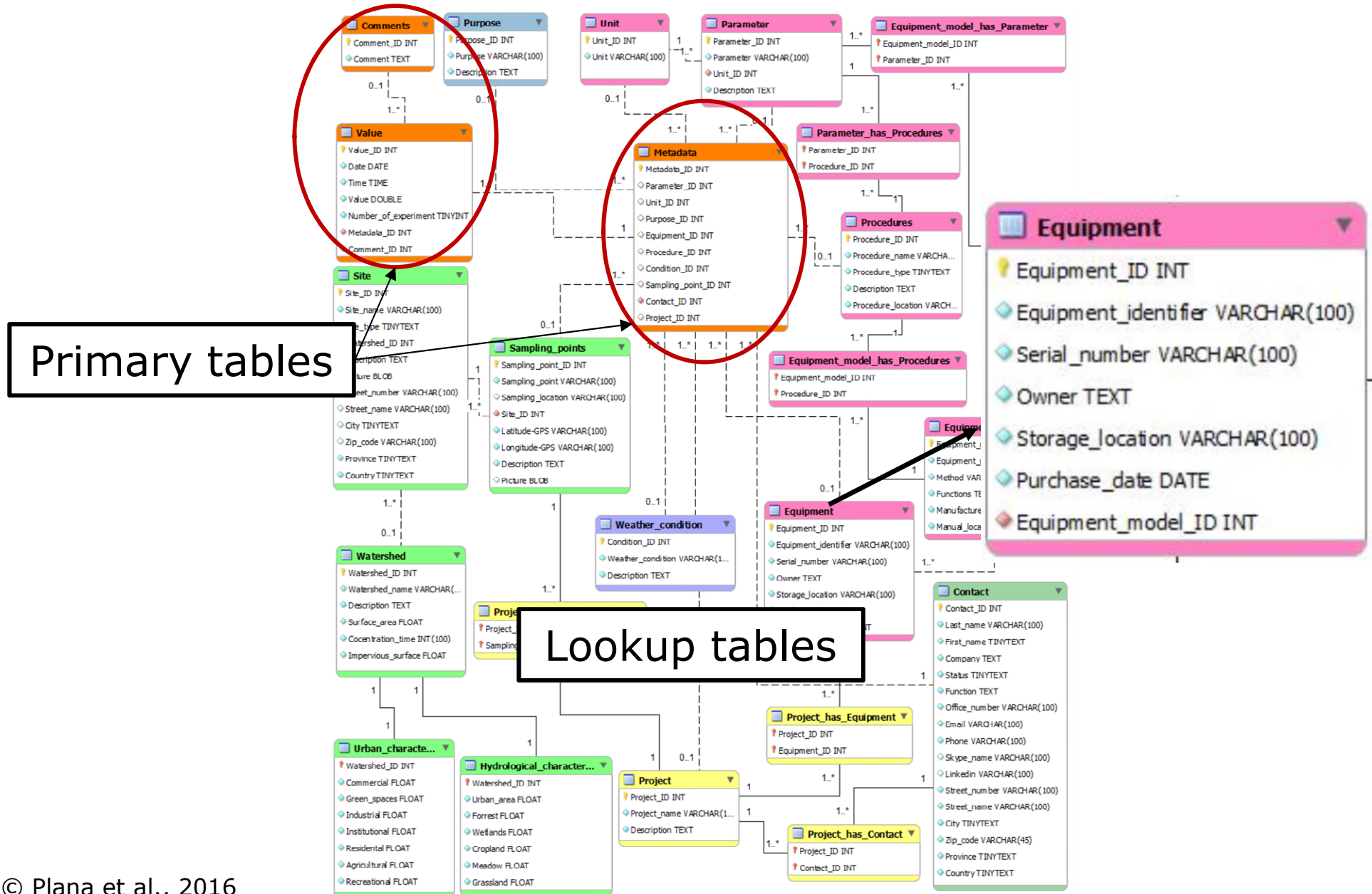
dat*EAU*base design – Primary tables
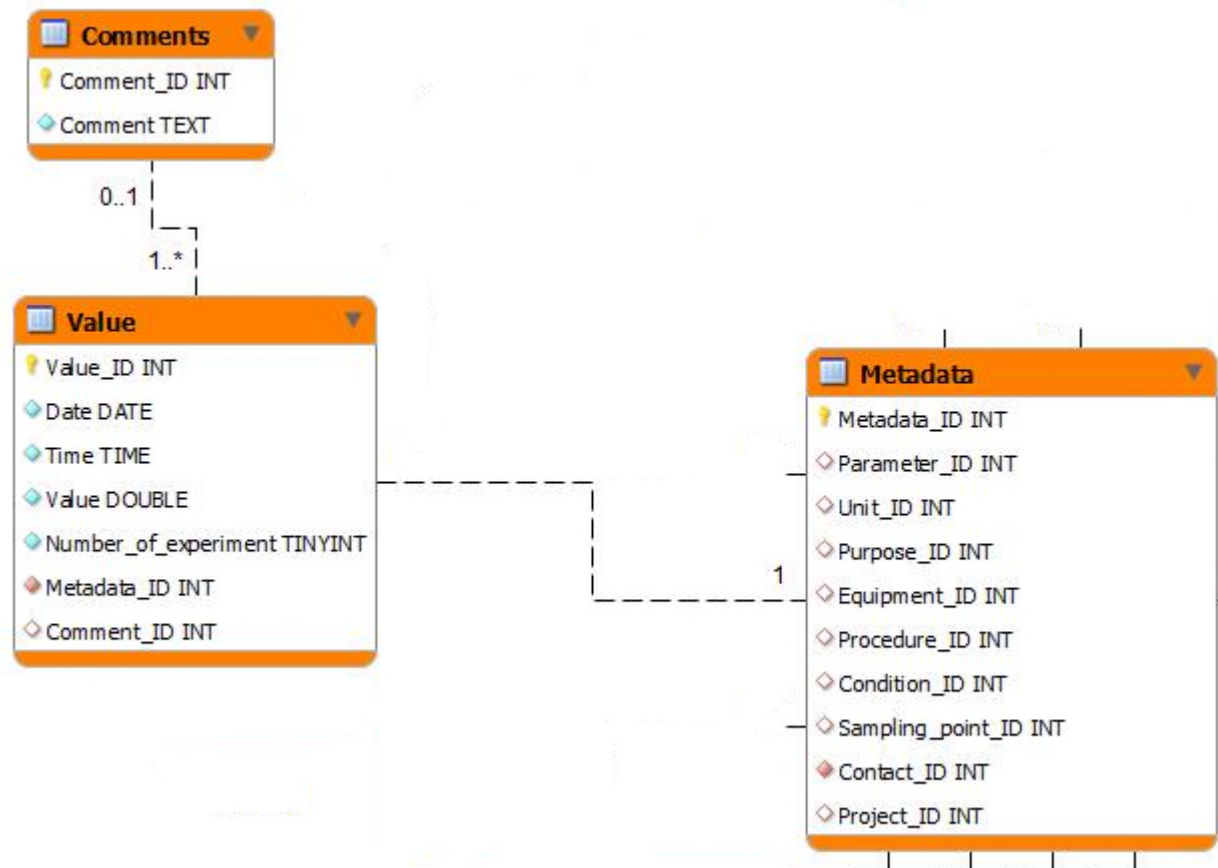
dat*EAU*base design – Primary tables

- **Date:** May 29, 2015

- **Time:** 15:36:02 GMT

- **Value:** 7.32

- **Parameter:** pH

- **Units:** [-]

- **Sensor:** pH_003

- **Conditions:** dry weather

- **Purpose:** calibration

- **Procedure:** ISO-15839

- **Site:** Grandes-Piles F/AL

- **Sampling point:** inlet

- **Responsible:** Plana

- **Project:** mon*EAU*

- **Comment:** Unsuccessful calibration

dat*EAU*base design – Equipment & procedures
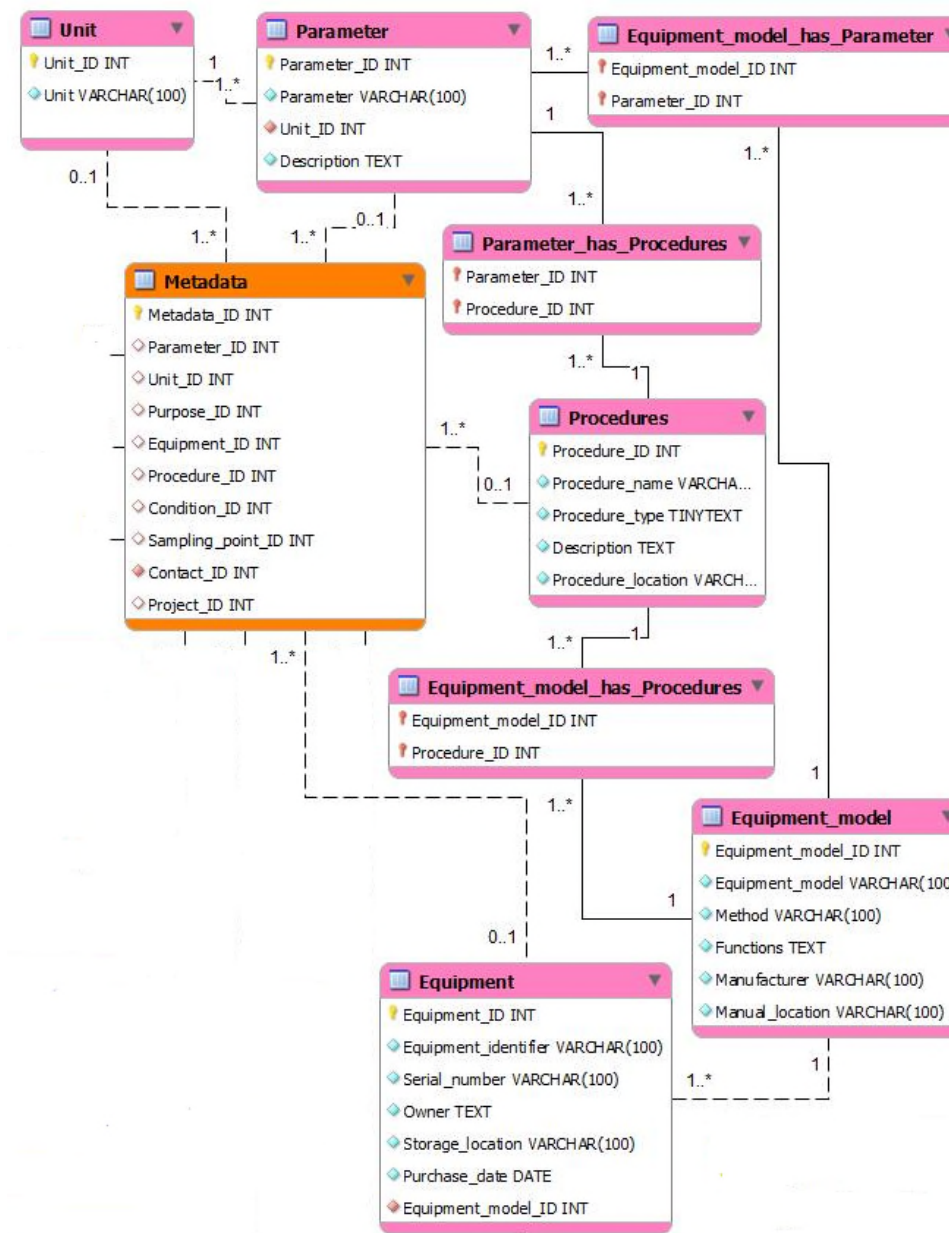
dat*EAU*base design – Equipment & procedures

- **Parameter:** pH

- **Units:** [-]

- **Sensor:** pH_003

- **Brand:** Hach

- **Model:** DPD1P1

- **Serial number:** 2659777

- **Principle:** Differential of electrical potential

- **Current location:** Grandes-Piles F/AL

- **Manual:** pHD sc Digital Differential Sensor. User manual

- **Manual location:** http://modelEAU...

dat*EAU*base design – Sampling points

Metadata
- Metadata_ID INT
- Parameter_ID INT
- Unit_ID INT
- Purpose_ID INT
- Equipment_ID INT
- Procedure_ID INT
- Condition_ID INT
- Sampling_point_ID INT
- Contact_ID INT
- Project_ID INT

Site
- Site_ID INT
- Site_name VARCHAR(100)
- Site_type TINYTEXT
- Watershed_ID INT
- Description TEXT
- Picture BLOB
- Street_number VARCHAR(100)
- Street_name VARCHAR(100)
- City TINYTEXT
- Zip_code VARCHAR(100)
- Province TINYTEXT
- Country TINYTEXT

Sampling_points
- Sampling_point_ID INT
- Sampling_point VARCHAR(100)
- Sampling_location VARCHAR(100)
- Site_ID INT
- Latitude-GPS VARCHAR(100)
- Longitude-GPS VARCHAR(100)
- Description TEXT
- Picture BLOB

Watershed
- Watershed_ID INT
- Watershed_name VARCHAR(...
- Description TEXT
- Surface_area FLOAT
- Cocentration_time INT(100)
- Impervious_surface FLOAT

Urban_characte...
- Watershed_ID INT
- Commercial FLOAT
- Green_spaces FLOAT
- Industrial FLOAT
- Institutional FLOAT
- Residental FLOAT
- Agricultural FLOAT
- Recreational FLOAT

Hydrological_character...
- Watershed_ID INT
- Urban_area FLOAT
- Forrest FLOAT
- Wetlands FLOAT
- Cropland FLOAT
- Meadow FLOAT
- Grassland FLOAT

dat*EAU*base design – Sampling points

- **Site:** Grandes-Piles F/AL
- **Sampling point:** Inlet
- **Address:** 267-303 5e Av., Grandes-Piles, G0X 1H0, QC, Canada
- **Coordinates:** 46°41'04"N 72°42'59"W
- **Watershed:** Saint-Maurice river
- **Surface area:** 43 300 km^2
- **Concentration time:** 2 days
- **Impervious surface:** 4 %

- **Urban characteristics:**
  - 54.25 % of green spaces
  - 2.25 % of industrial area
  - 13.5 % of residential area
  - 22 % of agricultural area
  - 8 % of recreational area

dat*EAU*base design - Projects
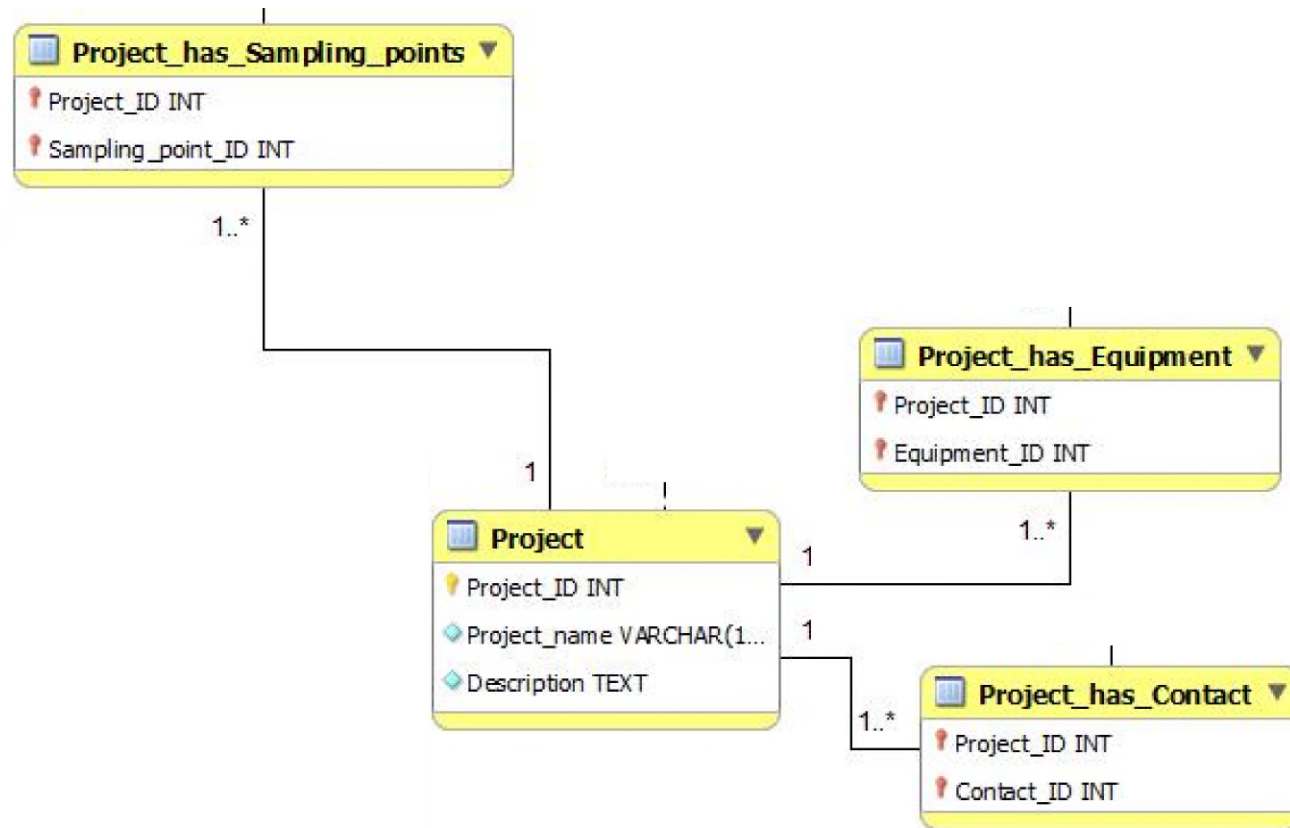
dat*EAU*base design - Projects

- **Project name:** mon*EAU*

- **Description:** AMS to study the water quality

- **Sampling point:** inlet in Grandes-Piles F/AL

- **Equipment:**
  - conductivity_001
  - pH_003
  - ammolyser_001

- **Personnel involved:**
  - Alferes
  - Plana
  - Vanrolleghem

dat*EAU*base design – Other lookup tables

- **Contact information:**
  - **First name:** Queralt
  - **Last name:** Plana
  - **Company:** Université Laval
  - **Status:** PhD student
  - **Address:** 1065, avenue de la Médecine, room PLT-2954, Québec, G1V 0A6, QC, Canada
  - **E-mail:** queralt.plana.1@ulaval.ca
  - **Phone:** +1 418 656 2131, ext. 8730

dat*EAU*base design – Other lookup tables

- **Weather conditions:**

  - **Condition:** wet day

  - **Description:** rainfall of more than 3 mm/d

- **Purpose:**

  - **Purpose:** sensor validation

  - **Description:** routine sensor validation activity for verification of proper operation

Contents

- Introduction

- Types of data

- Challenges

- dat*EAU*base scheme

- dat*EAU*base design

- **Examples of data**

- Conclusions

Example

- Understandable value

pH	May 29, 2_ 2 Dry weat_ _

Outlet of F/AL Grandes-Piles F/AL	7.2811	On-line measurement for water quality modelling

Plana	pH_002 sensor

Contents

- Introduction

- Types of data

- Challenges

- dat*EAU*base scheme

- dat*EAU*base design

- Examples of data

- **Conclusions**

Conclusions

- dat*EAU*base is a powerful tool to store data
  using a single consistent format,
  giving quality to the content

- Any environmental parameter can be stored into it

- It offers flexibility and it can be modified
  and adapted for future studies

- Its design is explicitly providing relevant metadata information
  to the measured values
  - The metadata is fundamental to understand the measured values
    for use in further studies

- It combines raw, filtered, validated, lab and metadata

Acknowledgments

Questions